

Von Suchmaschinen zu Antwortmaschinen: Semantische Technologien und Benutzerpartizipation im Web 3.0

Wolfgang Wahlster

Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI GmbH)
Saarbrücken, Kaiserslautern, Bremen und Berlin
<http://www.dfki.de/~wahlster/>

1. Einleitung

Die Suchmaschinen von morgen werden klar über die Leistungen der heute führenden Technologie von Google hinausgehen, indem sie nicht nur die Eingabe von Schlüsselwörtern sondern komplexe natürlichsprachlich formulierte Anfragen zulassen und Suchergebnisse höherer Qualität liefern, die zudem an den aktuellen Kontext und die jeweilige Aufgabenstellung des Benutzers angepasst sind. Sie werden Anfragen des Benutzers inhaltlich verstehen und statt einer Vielzahl von Verweisen auf Dokumente, in denen möglicherweise die gesuchte Information enthalten ist, gleich die Antwort auf die Benutzeranfrage liefern. So werden aus Suchmaschinen in Zukunft Antwortmaschinen, weil nicht mehr die Unterstützung eines Suchprozesses sondern die rasche Generierung einer gezielten Antwort auf eine konkrete Anfrage die Systemleistung ausmacht (vgl. Fig. 1).



Fig. 1: Stichwortbasierte Suche versus Fragebeantwortung

2. Von syntaktischen Suchmaschinen zu semantischen Antwortmaschinen

Besonders in mobilen Anwendungssituationen will der Benutzer als Ausgabe keine Listen von Verweisen auf möglicherweise relevante Webdokumente, die er dann wiederum weiter durchsuchen muss, sondern eine konzise Antwort auf seine Anfrage. So will der Familienvater während einer Autofahrt, wenn das Kind auf dem Rücksitz plötzlich über starke Schmerzen klagt, auf seine Frage „Wo ist der nächste Kinderarzt?“ nur die Antwort „Dr. Pfeifer in der Marienstrasse 23“, so dass er als Folgeauftrag das System bitten kann „Dann zeige mir die Route und führe mich dort hin“, um eine Navigationsunterstützung zum nächsten Kinderarzt zu erhalten. In dem vom Bundesministerium für Bildung und Forschung (BMBF) von 2004 – 2007 mit 13,7 Mio. Euro geförderten Verbundprojekt SmartWeb wurden unter meiner Leitung die Grundlagen für semantische Antwortmaschinen erarbeitet und erstmals eine solche Funktionalität für den mobilen Benutzer vollständig realisiert. In einem Mercedes-Testfahrzeug konnten sich die Besucher der CeBIT 2007 davon überzeugen, dass solche mobilen Antwortmaschinen heute schon in der Lage sind, erfolgreich Dialoge wie die als Beispiel angeführte Suche nach einem Kinderarzt mit dem Fahrer durchzuführen.

SmartWeb [14] zeigt damit den Weg auf, wie folgende Defizite bisheriger Suchmaschinen überwunden werden können:

- die Anfrage des Benutzers wird nicht inhaltlich verstanden, sondern nur als Zeichenkette betrachtet.
- jede Anfrage wird isoliert beantwortet, ohne den Inhalt vorangegangener Anfragen oder Suchergebnisse und damit den Dialogkontext zu berücksichtigen.
- Suchanfragen, die sich auf mehrstellige Relationen zwischen Begriffen beziehen und komplexe Restriktionen enthalten, führen zu unbefriedigenden oder sogar falschen Resultaten.
- „Versteckte Inhalte“, die im sog. „Deep Web“ nur über Portale, Webdienste oder in PDF-Dokumenten zugänglich sind, werden nicht gefunden.

Nutzer traditioneller Web-Suchmaschinen klagen wegen dieser Defizite oftmals über irrelevante Ergebnisse und einen zu umfangreichen Ergebnisraum. Sie verlassen die Suchmaschine ohne eine Antwort auf ihre Ursprungsfrage erhalten zu haben und versuchen weitere Recherchen mit anderen Suchmaschinen. SmartWeb (vgl. Fig. 2) bietet erstmals in einem integrierten Gesamtsystem alle Voraussetzungen, um diese Defizite zu überwinden und so von traditionellen Suchmaschinen zu mobilen Antwortmaschinen zu kommen.



Fig. 2: SmartWeb als mobile Antwortmaschine über die FIFA WM 2006

Ein anschauliches Beispiel für die Leistungsfähigkeit von SmartWeb als Antwortmaschine entstand bei einem Blindtest durch den für die Förderung von SmartWeb verantwortlichen Leiter des Referats 524 des BMBF, Rainer Jansen. Obwohl SmartWeb vielfach in der Suchdomäne „Sport“ und speziell „FIFA Fußballweltmeisterschaft 2006“ getestet war (vgl. Fig. 2), stellte Herr Jansen eine Frage aus dem Kulturbereich (siehe Fig. 3), die komplexe Relationen (Name von Sänger/in in der Titelrolle von La Traviata) und zwei Restriktionen (bei den Salzburger Festspielen, in der Premiere) enthielt. Obwohl eine ähnliche Frage SmartWeb zuvor noch nie gestellt worden war, wurde zur Verblüffung auch der anwesenden Fachwissenschaftler diese schwierige Frage präzise und korrekt mit „Anna Netrebko“ beantwortet.

Fig. 3 zeigt auch ein Beispiel dafür, dass der SmartWeb-Nutzer eine bestimmte Modalität der Antwort anfordern kann, indem er am unteren Rand der Bildschirmmaske von SmartWeb auf eines der vier Ikonen für Bild, Videoclip, Text- oder Tondokument mit dem Stift zeigt. Im vorliegenden Fall wird dann zusätzlich zu der Standardantwort in textueller Form noch ein Bild der Künstlerin ausgegeben, das durch eine Bildsuche im Web gefunden wurde.

Über alle Systemversionen hinweg bietet SmartWeb seinem Benutzer einen symmetrischen multimodalen Zugriff [11], der auf der Eingabeseite neben gesprochener und geschriebener Sprache, Gestik, Haptik und Video kombiniert mit Sprache, Haptik, Graphik, Video und Ton auf der Ausgabeseite. Der Benutzer bekommt die Informationen also nicht nur durch Eintippen von Fragen auf der kleinen Tastatur des mobilen Endgeräts, sondern auch durch gesprochene Sprache z.B. kombiniert mit Stifteingaben auf dem Bildschirm. Es wird also möglich, das Mobiltelefon in Alltagssprache zu ´fragen´ und dieses wird mit Informationen aus dem Netz ´antworten´. SmartWeb baut auf das BMBF-Vorgängerprojekt SmartKom (1999 – 2003) auf, das erstmals das Prinzip der symmetrischen Multimodalität beim semantischen In-

formationszugriff realisierte [10], allerdings im Gegensatz zu SmartWeb keinen freien Zugang zum offenen Web erlaubte, sondern ausschließlich in vorher festgelegten Diskursbereichen funktionierte.



Fig. 3: Beantwortung einer komplexen Suchanfrage im offenen Web durch SmartWeb

SmartWeb hat mit 16 Partnern aus Industrie und Forschung wesentliche wissenschaftliche und technische Grundlagen für das neue Leuchtturmprojekt *Theseus* gelegt, das im Rahmen der Hightech-Strategie der Bundesregierung vom Bundesministerium für Wirtschaft und Technologie (BMWi) zunächst mit 90 Mio. Euro ab 2007 für fünf Jahre gefördert wird. In *Theseus* sollen von 30 Konsortialpartnern die semantischen Technologien für innovative Anwendungen des zukünftigen Internet der Dienste erschlossen werden.

Das Internet der Dienste wird die automatische Komposition von Webdiensten ermöglichen, die zur Beantwortung komplexer Benutzeranfragen notwendig wird. Fragt beispielsweise eine Sekretärin während der Vorbereitung der nächsten Geschäftsreise ihres Chefs nach Japan „Wieviel Euro kostet die schnellste und preiswerteste Verbindung vom Flughafen Narita zum Palace Hotel in Tokio?“ dann sollte sie zukünftig auf der Basis der *Theseus*-Technologie die Antwort erhalten „Mit dem Narita-Express kostet dies derzeit 16,60 Euro“. Dazu muss die semantische Antwortmaschine nicht nur die Preise verschiedener Bahn- und Busverbindung und deren Fahrzeiten auf deren jeweiligen Webportalen finden, extrahieren und vergleichen, sondern auch die Kosten von 2700 Yen in Euro umrechnen. Dazu muss ein passender Webdienst wie der Universal Currency Converter automatisch gefunden und mit den entsprechenden Eingabeparametern versorgt werden, so dass nach dem aktuellen Umrechnungskurs die 16,60 Euro als Preis ausgegeben werden können. Wichtig ist auch die Dialogfähigkeit solcher Systeme, so dass eine Folgefrage wie „Wann gibt es den passenden An-

schluss zum Lufthansa-Flug nach Tokio“ im Kontext der Reiseplanung richtig verstanden wird, so dass „16.40, Gleis 4“ für den Narita-Express bestimmt wird. Da semantische Antwortmaschinen das gesamte Web als Wissensbasis nutzen, ist natürlich kritisch zu hinterfragen, ob man den Antworten vertrauen kann oder ob hier Fehlinformation aus unzuverlässigen Webseiten verarbeitet wurde. Daher müssen zukünftige Antwortmaschinen auch eine Erklärungskomponente enthalten, die es dem Benutzer erlaubt, die Informationsquellen und Webdienste kritisch zu prüfen, auf denen die Systemantwort beruht. In unserem Beispiel würde die Antwortmaschine den kritischen Nutzer darauf verweisen, dass die Information von der offiziellen und aktuellen Homepage der Japanischen Eisenbahnen stammt und der verwendete Währungsumrechner im Internet von einer seit 1993 zuverlässig arbeitenden Internetfirma XE.com betrieben wird.

3. Web 3.0 = Semantisches Web + Web 2.0

Das syntaktische Web 1.0 hat multimediale Dokumente und Dienste im Internet miteinander vernetzt, wobei die Relationen zwischen diesen Objekten keine maschinenverstehbare Semantik hatten, sondern lediglich zur hypermedialen Navigation des Benutzers dienten. Das Web 1.0 hat zwar den weltweiten Zugang zu digital gespeicherter Information drastisch verbessert, seine Inhalte sind aber nur maschinenlesbar, ohne maschinell verstehbar zu sein. Da Information im Web meist in natürlicher Sprache präsentiert wird, sind die gefundenen Dokumente bislang nur für den Menschen voll verständlich. Das semantische Web [3, 13] basiert dagegen auf der inhaltlichen Beschreibung digitaler Dokumente mit standardisierten Vokabularien, die eine maschinell verstehbare Semantik haben. Damit wird der Übergang von einem „Netz aus Verweisstrukturen“ zu einem „Netz aus Inhaltsstrukturen“ vollzogen. Den Kern der semantischen Technologie bilden Markierungssprachen, die eine formale Syntax und Semantik haben und in Form einer Ontologie eine standardisierte Begrifflichkeit zur Beschreibung digitaler Inhalte bereitstellen. Semantische Markierungssprachen wie OWL erlauben die weltweite Verbreitung und gemeinsame Nutzung von Ontologien. Die semantische Markierung durch Metadaten stellt neben der Markierung für das Layout (HTML) und die Struktur (XML) eines Dokumentes die dritte Ebene einer Dokumentbeschreibung dar. Bei der Suche nach Informationen können Schlussfolgerungen den Suchprozess beschleunigen oder gesuchte Information ableiten. Methoden des maschinellen Lernens, des Data Mining und des Text Mining reduzieren beim Aufbau und bei der Wartung von Ontologien den manuellen Aufwand. Software-Werkzeuge unterstützen das Erstellen, Editieren, Visualisieren, Auswerten, Verschmelzen und Validieren von Ontologien. Durch Zugriff auf semantische Webseiten können intelligente Suchagenten mit hoher Präzision gezielt Informationen

im Internet auffinden, filtern und kombinieren, um den Benutzer zeit-aufwändige Recherchearbeit abzunehmen.

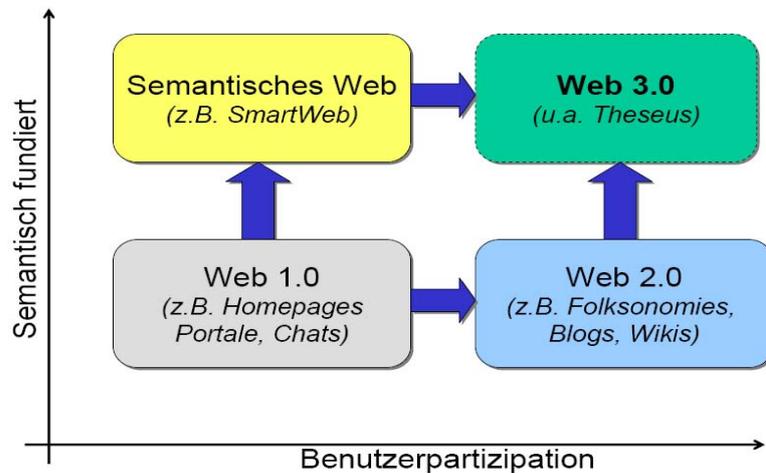


Fig. 4: Das Web 3.0 als Kombination von Semantischem Web und Web 2.0

Noch sind semantische Suchmaschinen nicht universell einsatzfähig, weil die Zahl der bereits semantisch annotierten Webdokumente und Webdienste noch im einstelligen Prozentbereich liegt. Dies ist auf den hohen manuellen Aufwand bei der nachträglichen semantischen Annotation bestehender Webseiten zurückzuführen. Sprachtechnologische Verfahren zur automatischen semantischen Annotation von Webseiten werden zwar immer leistungsfähiger, erreichen aber nur selten und dann in sehr speziellen Wissensbereichen die Qualität und Zuverlässigkeit menschlicher Annotationen.

Es liegt daher nahe, das Prinzip der Benutzerpartizipation des Web 2.0 als Massenbewegung für den Aufbau des semantischen Webs zu nutzen. Im Web 2.0 ist der Konsument gleichzeitig auch Produzent (ein *Prosumer*) und ist daher motiviert, die von ihm mitgestalteten Wikis, Photo-, Video- und Musiksammlungen stets zu aktualisieren und zu pflegen. Wenn es gelingt, auch den nichtprofessionellen Benutzer aktiv dabei mitwirken zu lassen, semantisch fundierte Webseiten und Webdienste zu schaffen, entsteht das Web 3.0 (vgl. Fig. 4). Dabei werden ontologische Klassifikationen in einfache Interaktionssequenzen auf natürliche Weise eingebunden, so dass es auch für den informatisch nicht geschulten Nutzer möglich wird, sein Weltwissen nutzbringend beim Aufbau des Web 3.0 einzubringen, um somit die Voraussetzung für hochpräzise Antwortmaschinen zu schaffen. Diesen Ansatz der Kombination aus semantischem Web und Web 2.0 zum neuartigen Web 3.0 der nächsten Generation verfolgen wir im Theseus-Konsortium.

4. SmartWeb: Eine multimodale und mobile Antwortmaschine

SmartWeb besteht aus drei großen Softwarebereichen [14, 7, 8]: den mobilen Subsystemen (für Mobiltelefone der MDA-Serie von T-Mobile, Autos und Motorräder), der multimodalen Dialogverarbeitung und den semantischen Zugriffsdiensten (vgl. Fig. 5). SmartWeb kann parallel eine Vielzahl von Benutzern unterstützen, weil das System nach dem Client-Server-Prinzip arbeitet und der Server gleichzeitig mehrere Dialoge mit verschiedenen mobilen Nutzern abwickeln kann. Dazu wird von der SmartWeb-Dialogplattform für jeden neuen eingehenden Anruf eine neue Instanz des Dialogservers instantiiert. Die Hauptverarbeitungsleistung von SmartWeb wird nicht auf dem mobilen Endgerät, sondern auf dem über eine breitbandige Funkverbindung erreichbaren Server erbracht, auf dem eine leistungsfähige Spracherkennung mit einer semantischen Anfrageanalyse und einer Komponente zum Dialogmanagement kombiniert ist. Auf dem MDA-Client läuft eine Java-basierte Steuerungskomponente für alle multimodalen Ein- und Ausgaben sowie die graphische Bedienoberfläche von SmartWeb. Falls einmal die Verbindung zum Sprachdialog-Server abbricht, verfügt der MDA zur Überbrückung der Störung auch über eine einfache Sprachbedienung auf der Basis von VoiceXML.

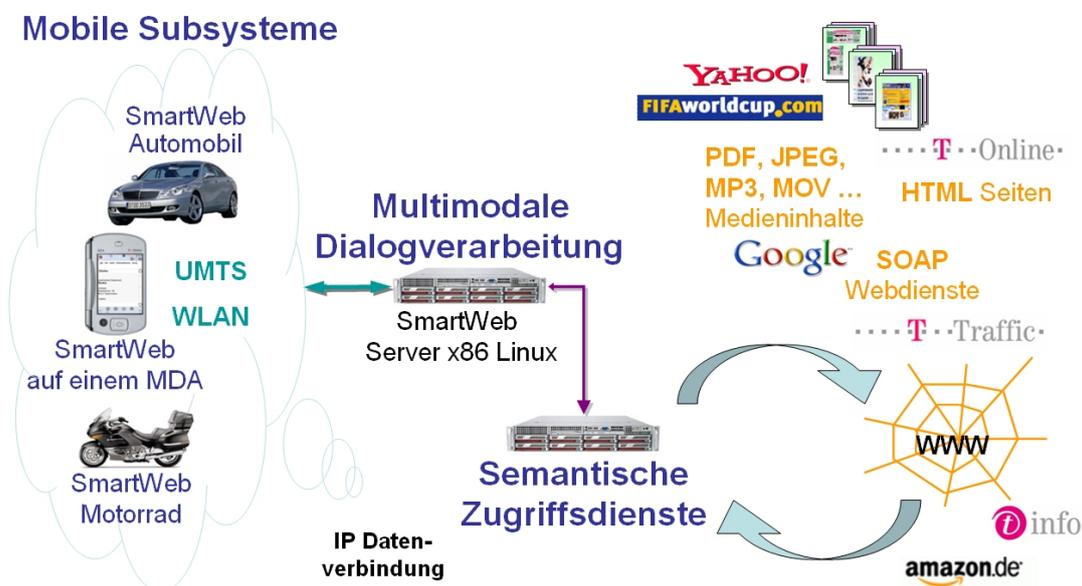


Fig. 5: Die drei großen Softwarebereiche von SmartWeb

Die Beantwortung von Anfragen an SmartWeb wird nicht nur über ein einziges Verfahren realisiert, sondern ein semantischer Mediator (vgl. Fig. 6) entscheidet, wie eine vom Dialogsystem analysierte Anfrage mithilfe der verschiedenen semantischen Zugriffsdienste von SmartWeb beantwortet werden soll. Neben der gezielten Informationsextraktion aus Passa-

gen von Dokumenten im offenen Web, die zunächst mit klassischen Suchmaschinen wie Google gefunden werden, kann z.B. auch eine Komposition von Webdiensten oder ein Aufruf eines Zugriffsagenten auf ein Portal durch den semantischen Mediator ausgelöst werden. Wie in Verbmobil [12] wird in SmartWeb also ein Ansatz verfolgt, der mehrere Lösungswege parallel vorhält und je nach Aufgabenstellung den besten Ansatz zur Reaktion auf die Benutzereingabe verfolgt.

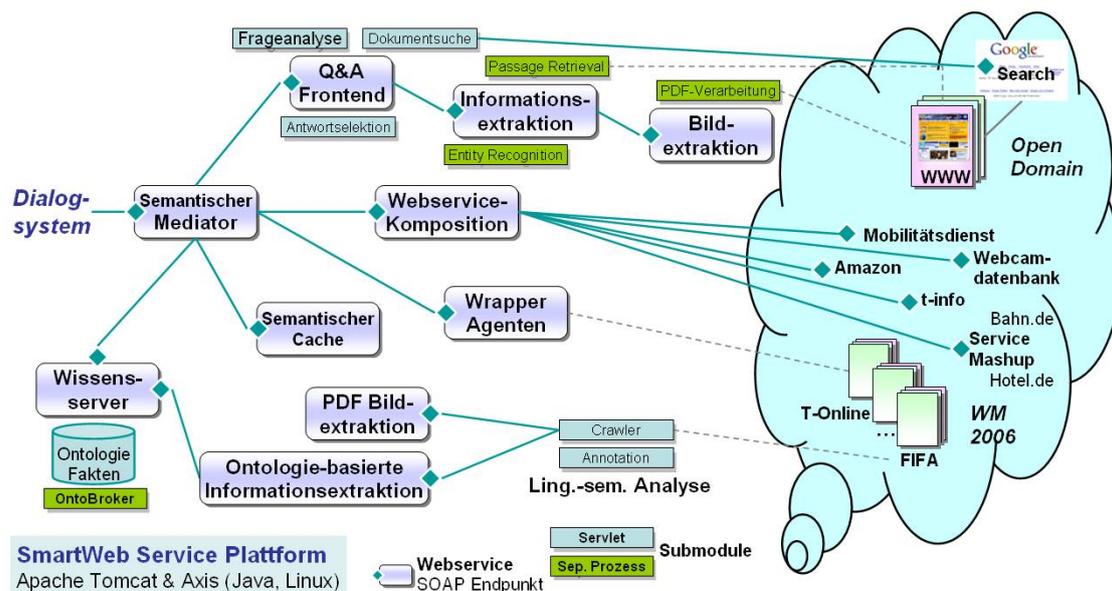


Fig. 6: Die Ansteuerung von Zugriffsdiensten durch den semantischen Mediator

SmartWeb kann neun verschiedene Webdienste automatisch je nach Bedarf zu höherwertigen Diensten komponieren (u.a. Routenplanung, Wetterinformation, Veranstaltungskalender, Verkehrsinformation, Webcam-Suche). Dabei nutzt das System die semantische Beschreibung dieser Webdienste mithilfe von Erweiterungen der standardisierten Dienstbeschreibungssprache OWL-S, die auf einer Beschreibungslogik mit effizienten Inferenzverfahren beruht. Die Komposition der Dienste erfolgt in SmartWeb über einen musterbasierten Planungsalgorithmus, der bei Unterspezifikation von Anfragen auch einen Klärungsdialog mit dem Benutzer initiieren kann (vgl. Fig. 7). Ein wichtiges Entwurfsprinzip von SmartWeb war die gezielte Benutzung von Ontologien in allen Verarbeitungskomponenten [6]. Die ontologische Infrastruktur SWIntO (**SmartWeb Integrated Ontology**) basiert auf einer Kombination der weit verbreiteten Universalontologien DOLCE und SUMO, die durch spezielle Diskurs- und Medienontologien ergänzt wurden [6]. Der W3C-Standard EMMA (**Extensible MultiModal Annotation Markup Language**) wurde in SmartWeb so ausgebaut, dass nicht nur die Semantik der multimodalen Benutzereingaben sondern auch der multimodalen Systemausgaben formal damit beschrieben werden kann, so dass alle Komponenten auf dem Server Information im EMMA-Format austauschen können.

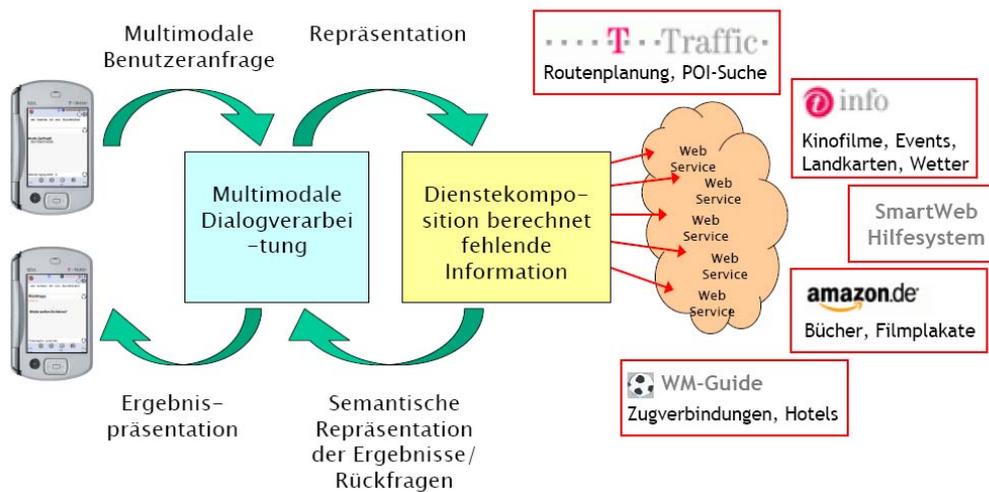


Fig. 7: Planbasierte Dienstekomposition in SmartWeb

Da wir heute noch ganz am Anfang des semantischen Web als zukünftiges Internet stehen und nur ganz wenige Webinhalte schon ontologisch annotiert sind, wurde in SmartWeb auch intensiv an der automatischen Generierung von semantischen Annotationen für bestehende HTML-Seiten, Photos und PDF-Dokumente aus dem Web gearbeitet. Anhand eines Korpus zum Bereich „Fußball“ wurden 5000 HTML-Seiten, 300 PDF-Dokumente und 15.000 Photos mit sprachtechnologischen Methoden der Informationsextraktion [2] sowie wissensbasierten Layout- und Bildanalyseverfahren bearbeitet. Für die Anwendungsdomäne „FIFA Fußballweltmeisterschaft 2006“ konnten im finalen Prototypen von SmartWeb 5000 Photos anhand der Bildunterschriften mit Spielen und Personen und 3400 Textpassagen zu Spielereignissen automatisch semantisch annotiert werden. Mit 110 komplett annotierten Mannschaftsphotos, 40 markierten Videosequenzen und 12 Audioreportagen ergeben sich rund 100.000 Ontologieinstanzen in einer umfangreichen Wissensbasis, die manuell erstellte mit automatisch extrahierten semantischen Annotationen aller Medienobjekte kombiniert.

5. Multimodale und lokationssensitive Antwortmaschinen

In SmartWeb kann die Handykamera der MDA-Telephone genutzt werden, um das Eintippen oder Sprechen komplexer Objektnamen bei Anfragen an das System zu umgehen. Der Benutzer photographiert einfach das Referenzobjekt selbst, ein Photo oder ein Modell davon und ein servergestütztes Objekterkennungssystem ordnet dann als Webdienst aufgrund charakteristischer Bildpunkte selbst bei geändertem Aufnahmewinkel und modifizierten Beleuchtungsparametern das Photo einem in einer Bilddatenbank abgelegten Referenzbild zu. Hier wird also erstmals eine Bildeingabe mit einer Spracheingabe kombiniert, so dass crossmoda-

le Fragen wie in Fig. 8 möglich werden: Nachdem der Nutzer ein Modell des Brandenburger Tores mit seiner Handykamera photographiert hat, kann er einfach fragen: „Wie komme ich dahin?“

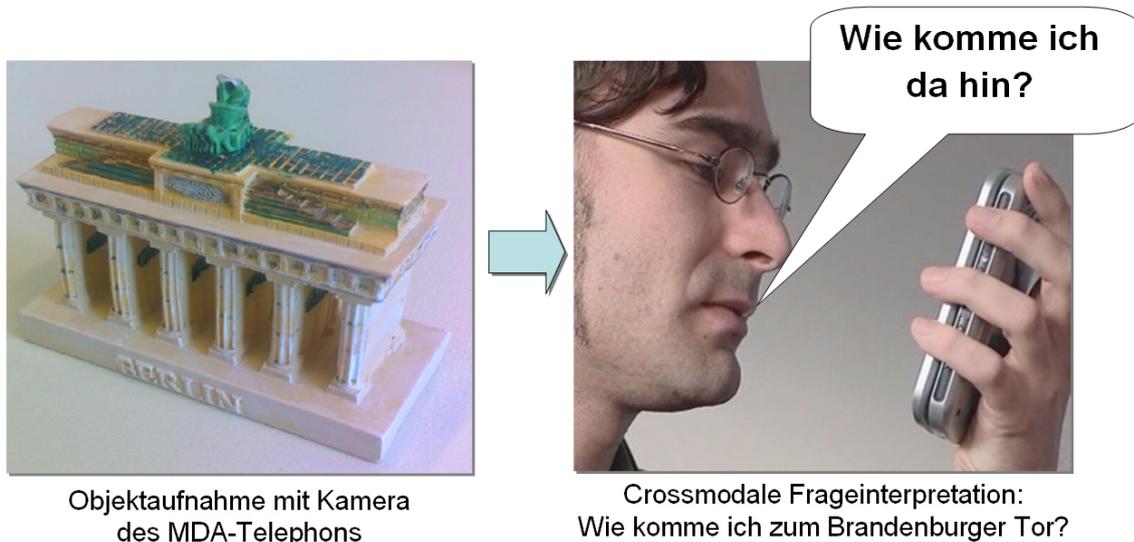


Fig. 8: Crossmodale Anfragen durch Kombination von Handykamera und Spracheingabe

Eine weitere Innovation von SmartWeb ist die Möglichkeit zur Beantwortung von cross-modalen Anfragen, die sich auf Bildmaterial beziehen, welches als Ergebnis einer vorangegangenen Anfrage von SmartWeb auf dem MDA präsentiert wurde. SmartWeb nutzt bei einer Folgefrage wie „Wer ist der dritte links oben?“ (vgl. Fig. 9) dazu die MPEG7-Repräsentation der Bildsegmente und interpretiert räumliche Relationen wie den Ausgangspunkt (links oben), um den referenzierten Spieler (der dritte) zu bestimmen.



Fig. 9: Beantwortung von Anfragen mit crossmodaler räumlicher Referenz auf Bildmaterial

Der Spieler mit der Nummer 18 wird durch die semantische Annotation des Bildes als „Tim Borowski“ bestimmt, so dass die Antwort von SmartWeb ausgegeben werden kann. Die Bilder sind dazu dekomponiert und die einzelnen Bildsegmente werden mit Instanzen aus der Medienontologie annotiert, die wiederum mit einer Sportontologie verknüpft ist, in der z.B. „Lehmann“ als Torwart klassifiziert ist.

Eine Weltneuheit bietet SmartWeb auch mit Sprachdialogen, in denen das System mehrere Internet-Dienste mit Fahrzeugdiensten automatisch verknüpft: Nachdem der Fahrer auf die lokationssensitive Frage „Wo gibt es hier Italiener?“ entsprechende Restaurants auf der digitalen Karte angezeigt bekommt und eines ausgesucht hat, kann er den frei formulierten Dialog fortsetzen mit „Wie komme ich dahin und vorher muss ich noch tanken“. Er bekommt dann eine Routenplanung zum Restaurant, die auch noch einen Tankstopp einplant (vgl. Fig. 10).

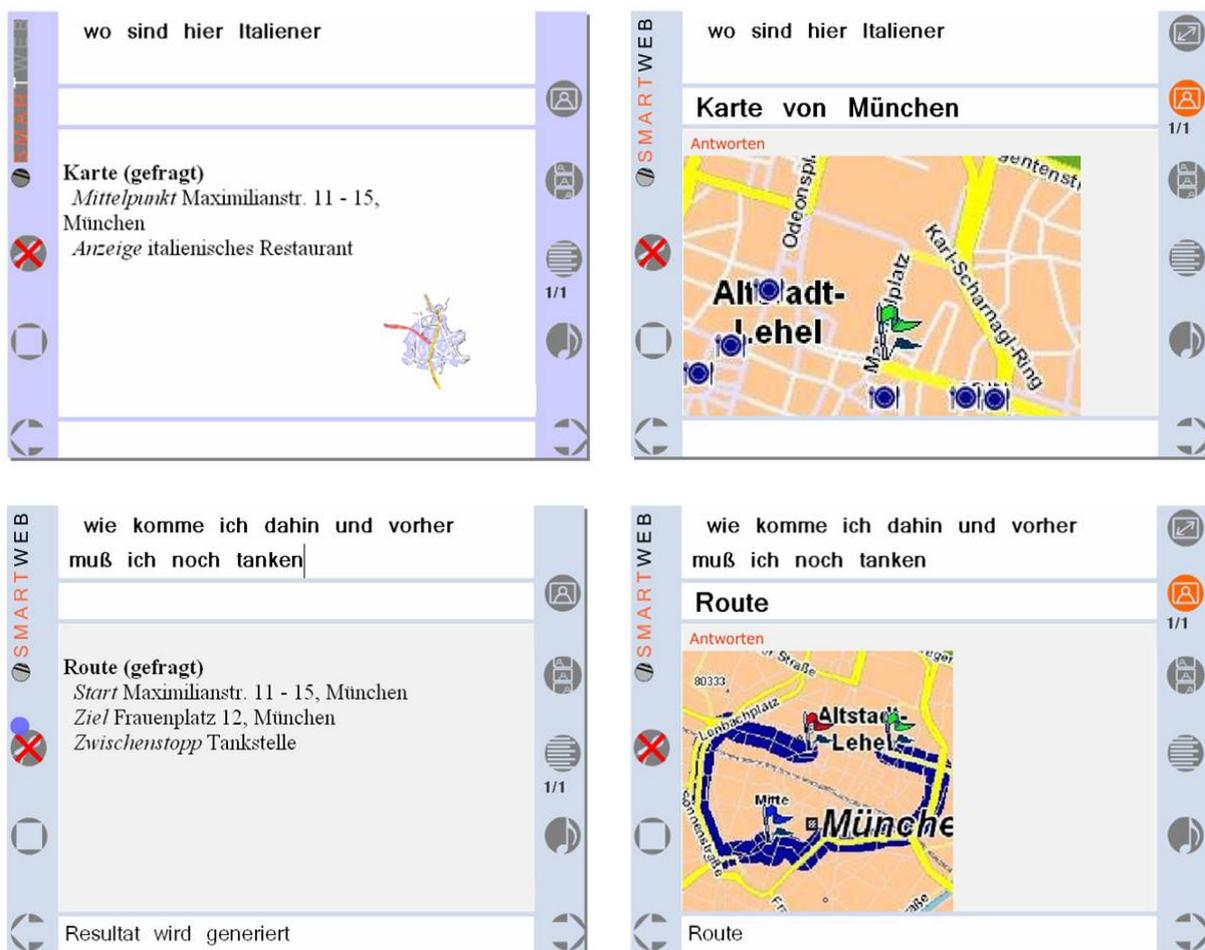


Fig. 10: Kohärenter Navigationsdialog durch die Kombination von mehreren Webdiensten

6. Semantische Videosuche und Peer-to-Peer Suchmaschinen

Das Aufkommen von digitalen Videokameras, Online-Videoportalen wie YouTube sowie digitaler Fernsehempfang über DVB-T ermöglicht es Anwendern, schnell große Mengen digitalen Videomaterials zu erzeugen, zu speichern und im Web zu veröffentlichen. Im Projekt InFiRe [9] entwickelten Forscher des DFKI unter Leitung von Thomas Breuel ein System für die inhaltsbasierte Suche in großen Videodatenbanken. Die Suche von InFiRe erfolgt dabei nach visuellen Eigenschaften des Videomaterials. Mit der Auswahl eines Keyframes aus der Videodatenbank wird die Suche nach ähnlichen Videos gestartet (vgl. Fig.11).

InFiRe

score = 1.0
ZDF/ZDF-07-04-07-0087.flv

score = 0.501228
ZDF/ZDF-07-04-06-0033.flv

score = 0.308033
ARD/ARD-07-04-05-0067.flv

score = 0.274157
ARD/ARD-07-04-05-0068.flv

score = 0.255269
ZDF/ZDF-07-04-07-0089.flv

score = 0.241668
ZDF/ZDF-07-04-06-0029.flv

more results requery

Random Images - Click one to start a query

Fig. 11: Semantische Videosuche mit InFiRe

Der Vergleich des Datenbankinhalts zur Anfrage erfolgt über eine Reihe von visuellen Merkmalen wie Farbe, Textur oder Bewegung. Durch Variation der verwendeten Merkmale und deren Gewichtung zueinander kann der Anwender Einfluss auf die Eigenschaften des Such-

ergebnisses nehmen. Eine Gewichtung einzelner Suchergebnisse hinsichtlich ihrer Relevanz erlaubt eine weitere Verbesserung des Resultats. Eine Demonstration kann unter <http://demos.iupr.org/cgi-bin/infire.cgi> aufgerufen werden. Bei der Verwendung spezieller Videomerkmale und Distanzmaße lässt sich auch eine explizite Suche nach identischen Videoinhalten durchführen, was z.B. zum Auffinden von urheberrechtlich geschütztem Material in Videodatenbanken genutzt werden kann. Momentan liegt der Schwerpunkt der Entwicklung auf der automatischen ontologischen Annotation von Videodaten aus Online-Videoportalen. Das System lernt dabei die visuellen Eigenschaften der Videos zu einem Schlagwort (z.B. Fußball, Schwimmen, Interview, Talkshow) autonom, indem es automatisch Videos aus den Online-Portalen herunterlädt und deren Merkmale lernt. Die Annotationen können dann automatisch einem Benutzer zur Auswahl vorgeschlagen werden (<http://demos.iupr.org/cgi-bin/videotagging.cgi>). In Kombination mit anderen Merkmalen können sie auch wiederum für das semantische Video-Retrieval eingesetzt werden.

Die Peer-to-Peer-Technologie hat das Potential, das Monopol der Webindexierung durch heutige Suchmaschinen-Firmen zu brechen. Dabei wird eine benutzerdedizierte, semantische und kontextsensitive Suchmaschine auf jedem Peer-Rechner installiert und soziale Netze werden bei der Kollaboration der Peers systematisch genutzt (vgl. Fig. 12). Über verteilte Hash-Tabellen entstehen skalierbare und selbstorganisierende Datenstrukturen und Suchverfahren über Millionen von Peers.

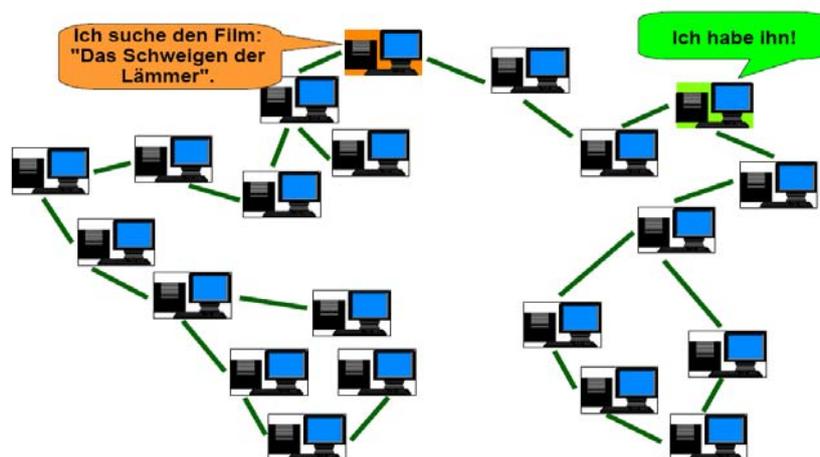


Fig. 12: Semantische Peer-to-Peer Suche

Dabei werden lokal berechnete PageRank-Werte für die Relevanz von Dokumenten aggregiert und bei zufälligen Rendezvous mit anderen Peers ausgetauscht [1]. Es entsteht ein dezentraler Index mit lokal berechneten Autoritätswerten für die Güte eines Dokumentes, der gegen die global berechneten Werte konventioneller Suchmaschinen konvergiert, ohne dabei den hohen Aufwand zentraler Datenzentren zu erfordern.

7. Konklusion

In Zukunft geht die Entwicklung von Suchmaschinen hin zu Antwortmaschinen. Semantische Technologien auf der Basis maschinenverstehbarer ontologischer Annotationen von Webinhalten sind die Basis für hochpräzise Suchagenten und mobile Antwortmaschinen. Das zukünftige Web 3.0 wird aus einer Kombination der Ansätze des semantischen Webs und des Web 2.0 entstehen. Dabei wird das Prinzip der Benutzerpartizipation des Web 2.0 mit der ontologisch fundierten Beschreibung semantischer Webinhalte kombiniert. Komplexe Anfragen können nur durch die automatische Komposition semantischer Webdienste beantwortet werden, so dass ein Internet der Dienste entstehen muss. Die Semantische Videosuche wird verstärkt auf maschinellen Lernverfahren beruhen, die autonom eine ontologische Annotation aus Bildmerkmalen ableiten. Durch innovative Peer-to-Peer-Ansätze könnte langfristig das heute Indexierungsmonopol der Suchmaschinen-Firmen gebrochen werden und durch die Nutzung von Metawissen über die jeweilige Expertise der Peer-Nutzer eine weitere Steigerung der Antwortpräzision erreicht werden.

8. Danksagung

Als Gesamtprojektleiter danke ich dem Bundesministerium für Bildung und Forschung (BMBF) für die Förderung von SmartWeb von 2004 – 2007 unter dem Förderkennzeichen 01 IMD 01 sowie allen beteiligten Industriepartnern für die zusätzlich eingesetzten Investitionsmittel. Insgesamt ist SmartWeb eine Teamleistung von über 100 Wissenschaftlern und Softwareingenieuren, die in einer verteilten Entwicklungsumgebung ein multimodales Dialogsystem geschaffen haben, das trotz seiner weltweit einzigartigen Leistungsfähigkeit eine große Robustheit unter Realzeitbedingungen aufweist. Daher möchte ich mich an dieser Stelle bei allen Mitarbeitern des SmartWeb-Konsortiums sehr herzlich für die exzellente Kooperation bedanken.

9. Literatur

- [1] Bender, Matthias, Michel, Sebastian, Triantafillou, Peter, Weikum, Gerhard, Zimmer, Christian (2006): P2P Content Search: Give the Web Back to the People. In: 5th International Workshop on Peer-to-Peer Systems (IPTPS 2006), Santa Barbara, USA, 2006.
- [2] Cramer, Irene, Leidner, Jochen, Klakow, Dietrich (2006): Building an Evaluation Corpus for German Question Answering by Harvesting Wikipedia. In: Proceedings of LREC 2006, International Conference on Language Resources and Evaluation, Mai 2006, Genova, Italy
- [3] Fensel, Dieter, Hendler, Jim, Lieberman, Henry, Wahlster, Wolfgang (2003): Spinning the Semantic Web. Bringing the World Wide Web to Its Full Potential. Cambridge: MIT Press 2003.

- [4] Kaiser, Moritz, Mögele, Hans, Schiel, Florian (2006): Bikers Accessing the Web: The SmartWeb Motorbike Corpus. In: Proceedings of the LREC 2006, International Conference on Language Resources and Evaluation, Juni 2006, Genova, Italy
- [5] Maybury, Mark, Wahlster, Wolfgang (eds.) (1998): Readings in Intelligent User Interfaces. San Francisco: Morgan Kaufmann, 1998.
- [6] Oberle, Daniel, Ankolekar, Anuypriya, Hitzler, Pascal, Cimiano, Philipp, Sintek, Michael, Kiesel, Malte, Mougouie, Babak, Vembu, Shankar, Baumann, Stephan, Romanelli, Massimo, Buitelaar, Paul, Engel, Ralf, Sonntag, Daniel, Reithinger, Norbert, Loos, Berenike, Porzel, Robert, Zorn, Hans-Peter, Micelli, Vanessa, Schmidt, Christian, Weiten, Moritz, Burkhardt, Felix, Zhou, Jianshen (2007): DOLCE ergo SUMO: On Foundational and Domain Models in SWIntO (SmartWeb Integrated Ontology). Journal of Web Semantics, 3, 2007.
- [7] Reithinger, Norbert, Herzog, Gerd, Blocher, Anselm (2007): SmartWeb - Mobile Broadband Access to the Semantic Web. In: KI - Künstliche Intelligenz, 2/2007, p. 30-33.
- [8] Sonntag, Daniel, Engel, Ralf, Herzog, Gerd, Pfalzgraf, Alexander, Pflieger, Norbert Romanelli, Massimo, Reithinger, Norbert (2007): Smart Web Handheld - Multimodal Interaction with Ontological Knowledge Bases and Semantic Web Services. In: Huang, Thomas Nijholt, Anton, Pantic, Maja, Plentland, Alex (eds.): Artificial Intelligence for Human Computing. LNCS 4451, Berlin, Heidelberg, Springer, 2007, p. 272-295
- [9] Ulges, Adrean, Schulze, Christian, Keysers, Daniel, Breuel, Thomas (2007): Content-based Video Tagging for Online Portals. In: MUSCLE Workshop 2007, Third MUSCLE/ImageCLEF Workshop on Image and Video Retrieval Evaluation, Budapest, Hungary, September 2007.
- [10] Wahlster, Wolfgang (ed.) (2006): SmartKom: Foundations of Multimodal Dialogue Systems. Cognitive Technologies Series, Heidelberg, Germany: Springer, 2006, 644 pp.
- [11] Wahlster, Wolfgang (2003): Towards Symmetric Multimodality: Fusion and Fission of Speech, Gesture, and Facial Expression. In: Günter, A., Kruse, R., Neumann, B. (eds.): KI 2003: Advances in Artificial Intelligence. Proceedings of the 26th German Conference on Artificial Intelligence, September 2003, Hamburg, Germany, Berlin, Heidelberg: Springer, LNAI 2821, 2003, p. 1 - 18
- [12] Wahlster, Wolfgang (2000) (ed.): Verbmobil: Foundations of Speech-to-Speech Translation. Berlin, Heidelberg, New York, Barcelona, Hong Kong, London, Milan, Paris, Singapore, Tokyo: July 2000, Springer.
- [13] Wahlster, Wolfgang (2004): Semantisches Web. In: Bullinger, H.-J. (ed.): Trendbarometer Technik: Visionäre Produkte, Neue Werkstoffe, Fabriken der Zukunft, München, Wien: Hanser, 2004, p. 62 - 63
- [14] Wahlster, Wolfgang (2007): SmartWeb: Ein multimodales Dialogsystem für das semantische Web. In: Reuse, B., Vollmar, R. (eds.): Informatikforschung in Deutschland, Heidelberg, Berlin, Springer, 2007